
Accessed from  http://eprints.qut.edu.au

Copyright 2006 Elsevier
Partially Observed Non-linear Risk-sensitive Optimal Stopping Control for Non-linear Discrete-time Systems

Jason J Ford

School of Engineering Systems, Queensland University of Technology, G.P.O. Box 2434 Brisbane 4001 Australia

Abstract

In this paper we introduce and solve the partially observed optimal stopping non-linear risk-sensitive stochastic control problem for discrete-time non-linear systems. The presented results are closely related to previous results for finite horizon partially observed risk-sensitive stochastic control problem. An information state approach is used and a new (three-way) separation principle established that leads to a forward dynamic programming equation and a backward dynamic programming inequality equation (both infinite dimensional). A verification theorem is given that establishes the optimal control and optimal stopping time. The risk-neutral optimal stopping stochastic control problem is also discussed.

Key words: Partially Observed, Optimal Stopping, Information State, Dynamic Programming, Stochastic Control.

Email address: j2.ford@qut.edu.au (Jason J Ford).

1 The author acknowledges that this work was conducted at and funded by the Weapons Systems Division, Defence Science Technology Organisation, Australia.

Preprint submitted to Elsevier Science 6 October 2005
1 Introduction

There has been much recent effort concerned with partially observed non-linear stochastic and dynamic games in both continuous and discrete time [8,7,3]. One important class of partially observed stochastic control problems are those with a risk-sensitive (or integral of exponential type) cost criterion. These risk-sensitive problems are important because under certain parameter limits both risk-neutral and dynamic game problems can often be recovered.

The solutions for the linear/quadratic risk sensitive stochastic control problem with incomplete state information in discrete-time and continuous-time were given in [11,12] and [3] respectively. In [7] a solution to the partially observed non-linear risk-neutral stochastic control problem in discrete-time was developed through the information state approach and an appropriate separation principle and dynamic programming equations given. This information state approach was applied to the partially observed non-linear risk-sensitive stochastic control problems in [8] where again a separation principle and dynamic programming equations are established (and connections with dynamics games and conditions for certainty equivalence given).

In the standard optimal control problem, the controller is designed to influence the state dynamics, but must operated over a prescribed time-horizon. If the controller is allowed to select the final time-horizon then the problem is known as an optimal stopping time and control problem (also sometimes known as a “leavable control” problem). Typically, in the optimal stopping control problem, the controller is allowed the choose the stopping time in a closed loop manner and does so to balance the cost of stopping at a particular
time instant against the cost of continuing. Although, it may seem that the
optimal stopping and control is only a special case of the general optimal con-
trol problem (perhaps by incorporating the stopping action into the control
space), posing within the standard formulation - if even possible - would hide
important aspects of the solution structure. For example, in a simpler con-
text, the fully observed optimal stopping control problem leads to variational
inequality type solutions that require particular analysis tools [1].

Even though partially observed variants of the optimal stopping control
problem seem quite important (including applications such as the pricing of
real/American stock options, target tracking and trajectory planning prob-
lems) little effort appears to have been devoted to these problems.

The nature of the contribution here is also highlighted by considering recent
publications in the related area of optimal switching control (optimal stopping
control is a specific case of these types of problems)[13,14]. The apparent state-
of-the-art solution to the optimal switching control problem consists of a two-
state optimisation approach involving a computationally expensive search over
candidate switching instants [14]. Using the ideas presented here, the optimal
timing of these events can be found in feedback form involving only slightly
more complexity than a standard non-linear optimal control problem.

Our motivation for considering the partially observed optimal stopping and
control problem stems from guidance and trajectory design problems that have
natural optimal stopping features and are not suited to fixed finite-horizon
formulations.

The first paper in the area of partially observed optimal stopping and control
appears to be [9], where a continuous-time partially observed optimal stopping
control problem is considered that involves the tracking control of Brownian motion using a bounded control term whose gain is partially observed through the state process. The authors point out that the problem conceptually involve three aspects: estimation, control and stopping. The simple nature of the problem considered in [9] allows for a certainty equivalence principle to be established and the mixed control/stopping problem to be separated into a pure optimal control problem and a pure optimal stopping problem. Such certainty equivalence results do not appear to hold in general and typically the estimation, control and stopping aspects of the problem must be considered together.

In this paper we extend the results of [8,7] to the discrete-time partially observed optimal stopping non-linear risk-sensitive stochastic control problem. A key difference here is that the optimal stopping and control problem requires a three-way separation principle. We developed dynamic programming equations in an infinite dimensional information state and establish an appropriate separation principle (between the estimation problem, the control problem and the stopping problem). The dynamic programming equations for the risk-neutral variant of the problem are also established.

This paper is organised as follows: In Section 2, the discrete-time partially observed non-linear risk-sensitive optimal stopping control problem is introduced and an appropriate information state is developed. In Section 3, dynamic programming equations and a verification theorem are established that demonstrate a separation principle exists for this optimal stopping control problem. In Section 4, the partially observed risk-neutral optimal control stopping problem is introduced followed by appropriate dynamic programming equations. Finally, in Section 6, some conclusions are given.
2 The Risk-Sensitive Optimal Stopping Problem

2.1 Dynamics

We shall consider discrete-time dynamics with additive noise present in both the state and observation processes. Initially, consider a probability space \((\Omega, \mathcal{F}, P^u)\) where \(\omega \in \Omega\), \(\mathcal{F}\) is a \(\sigma\)-algebra on \(\Omega\) and \(P^u(\omega) : \Omega \to \mathbb{R}\) is a probability map. We begin to introduce the risk-sensitive optimal stopping stochastic control problem by considering the following non-linear discrete-time system defined on this probability space:

\[
\begin{align*}
    x_{k+1} &= a(x_k, u_k) + w_k \\
    y_{k+1} &= b(x_k) + v_k
\end{align*}
\] (1)

Here \(x_k \in \mathbb{R}^N\) represents the state of the system and is not directly measured but it is assumed \(x_0\) has known density \(\rho(x) = (2\pi)^{-N/2} \exp(-1/2|x|^2)\). Although more complicated densities can be considered (that have a certain compactness properties), for simplicity of presentation we consider only gaussian densities. The state noise \(\{w_k\}\) is a \(\mathbb{R}^N\) valued i.i.d noise sequence with density \(\psi(w) = (2\pi \epsilon)^{-N/2} \exp(-1/2\epsilon|w|^2)\). The observation process \(y_k \in \mathbb{R}\) is measured and can be used to select control actions \(u_k\) (and to decide whether to continue or stop the process). The observation noise \(\{v_k\}\) is a real valued i.i.d noise sequence with density \(\phi(v) = (2\pi \epsilon)^{-N/2} \exp(-1/2\epsilon|v|^2)\). Further, we assume that \(a \in C^1(\mathbb{R}^N \times \mathbb{R}^M, \mathbb{R}^N)\) and \(b \in C(\mathbb{R}^N)\) are bounded and uniformly continuous.

Let \(\mathcal{G}_k\) and \(\mathcal{Y}_k\) denote the complete filtrations generated by \((x_0, \ldots, x_k, y_0, \ldots, y_k)\) and \((y_0, \ldots, y_k)\) respectively. The controls \(u_k\) take values in \(U \subset \mathbb{R}^M\), assumed
compact, and are $\mathcal{Y}_k$ measurable. We write $\mathcal{U}_{k,\ell}$ for the set of admissible control processes defined on the interval $[k, \ell]$.

Let us assume a discrete-time horizon $[0, T]$ where $T > 0$ is the maximum allowable stopping time. Let $\tau$ denote a stopping time $\tau : \Omega \rightarrow [0, T]$ such that $\{\omega : \tau(\omega) \leq k\} \in \mathcal{Y}_k$ for all $k$ (that is, $\tau$ is $\mathcal{Y}_k$ measurable). See [4, Page 133] for a discussion of stopping times. Throughout this paper we assume that (1) describes the dynamics on the discrete-time interval $[0, \tau]$ where $\tau \leq T$.

We introduce an equivalent probability measure $\bar{P}$ under which $\{y_k\}$ is i.i.d. with density $\phi$, independent of $\{x_k\}$ and $x$ satisfies the first equation of (1). For $\tau \in [0, T]$ and $u \in \mathcal{U}_{0, \tau-1}$, the original probability measure can be defined in terms of this equivalent measure by setting the restriction of the Radon-Nikodym derivative to $\mathcal{G}_k$ equal to $Z_k$ as follows [6]:

$$
\left. \frac{dP^u}{dP} \right|_{\mathcal{G}_k} = Z_k = \prod_{\ell=1}^{k} \Psi(x_{\ell-1}, y_{\ell}) \quad \text{where} \quad \Psi(x, y) = \frac{\psi(y - b(x))}{\psi(y)}.
$$

2.2 Cost

The cost function for the optimal stopping risk-sensitive control problem for admissible $\tau \in [0, T]$ and $u \in \mathcal{U}_{0, \tau-1}$ is defined by

$$
J(u, \tau) = E^u \left[ \exp \frac{\mu}{\epsilon} \left( \sum_{\ell=0}^{\tau-1} L(x_{\ell}, u_{\ell}) + \Phi(x_{\tau}) \right) \right]
$$

and the partially observed optimal stopping risk-sensitive stochastic control problem is to find $\tau^* \in [0, T]$ and $u^* \in \mathcal{U}_{0, \tau^*-1}$ such that

$$
J(u^*, \tau^*) = \min_{\tau \in [0, T]} \inf_{u \in \mathcal{U}_{0, \tau-1}} J(u, \tau)
$$
To ensure the control problem is well posed, and to allow use of the results of [8], we make the following assumptions. These assumptions can be relaxed in various ways, for example by linear growth type assumptions, but these relaxations would make the presentation here more complicated (see [2] for some discussion of alternative formulations of the discrete-time non-linear optimal control problem).

We assume:

1. \( L \in C(R^N \times R^M) \) is nonnegative, bounded, and uniform continuous.
2. \( \Phi \in C(R^N) \) is nonnegative, bounded, and uniform continuous.

The cost can be rewritten in terms of the reference probability measure as

\[
J(u, \tau) = \bar{E} \left[ Z_{\tau} \exp \left( \frac{\mu}{\epsilon} \sum_{t=0}^{\tau-1} L(x_t, u_t) + \Phi(x_{\tau}) \right) \right]
\]

(3)

**Remark 1** The parameter \( \mu > 0 \) is the risk-sensitive parameter. Under our assumptions, the cost function is finite for all \( \mu > 0 \).

### 2.3 Information State

We consider the space \( L^\infty(R^N) \) and its dual \( L^{\infty\ast}(R^N) \), which includes \( L^1(R^N) \); see [10] for an introduction into vector space concepts. Let us introduce the \( \langle ., . \rangle \) notation to denote the operation of \( \nu \in L^1(R^N) \) and \( \eta \in L^\infty(R^N) \) as

\[
\langle \nu, \eta \rangle = \int_{R^N} \nu(x) \eta(x) dx.
\]

We now define an information state process \( \sigma_k \in L^{\infty\ast}(R^N) \) by

\[
\langle \sigma_k, \eta \rangle = \bar{E} \left[ \eta(x_k) \exp \left( \frac{\mu}{\epsilon} \sum_{t=0}^{\tau-1} L(x_t, u_t) \right) Z_k \bigg| Y_k \right]
\]

(4)
for all test functions \( \eta \) in \( L^\infty(\mathbb{R}^N) \), for \( k = 1, \ldots, \tau \) and \( \sigma_0 = \rho \in L^1(\mathbb{R}^N) \). We introduce the bounded linear operator \( \Sigma : L^\infty(\mathbb{R}^N) \to L^\infty(\mathbb{R}^N) \) defined by

\[
\Sigma(u, y)\nu(\xi) = \int_{\mathbb{R}^N} \psi(z - a(\xi, u))\nu(z) \exp \left( \frac{\mu}{\epsilon} L(\xi, u) \right) \Psi(\xi, y).
\]

(5)

The bounded linear adjunct operator \( \Sigma^* : L^{\infty*}(\mathbb{R}^N) \to L^{\infty*}(\mathbb{R}^N) \) adjunct to \( \Sigma \) is defined by \( \langle \Sigma^* \zeta, \eta \rangle = \langle \zeta, \Sigma \eta \rangle \) for all \( \zeta \in L^{\infty*}(\mathbb{R}^N) \), \( \eta \in L^\infty(\mathbb{R}^N) \).

The following theorem establishes that \( \sigma_k \) is in \( L^1(\mathbb{R}^n) \) and its evolution is governed by the operator \( \Sigma^* \) given as follows

\[
\Sigma^*(u, y)\sigma(z) = \int_{\mathbb{R}^N} \psi(z - a(\xi, u)) \exp \left( \frac{\mu}{\epsilon} L(\xi, u) \right) \Psi(\xi, y)\sigma(\xi) d\xi.
\]

(6)

**Theorem 2** The information state \( \sigma_k^\rho \) satisfies the recursion

\[
\begin{cases}
\sigma_k^\rho = \Sigma^* (u_{k-1}, y_k) \sigma_{k-1}^\rho & \text{for } k \leq \tau \\
\sigma_0 = \rho.
\end{cases}
\]

(7)

Further, \( \sigma_k^\rho \in L^1(\mathbb{R}^N) \) since \( \rho \in L^1(\mathbb{R}^N) \) and \( \Sigma^* \) maps \( L^1(\mathbb{R}^N) \) into \( L^1(\mathbb{R}^N) \).

**PROOF.** Essentially the same result is proven in [8] following the steps of [7]. We repeat the proof here.

Consider for any test function \( \eta \in L^\infty(\mathbb{R}^N) \). Then

\[
\langle \sigma_k^\rho, \eta \rangle = \bar{E} \left[ \eta (a(x_{k-1}, u_{k-1}) + w_{k-1}) \exp \left( \frac{\mu}{\epsilon} L(x_{k-1}, u_{k-1}) \right) \Psi(x_{k-1}, u_k) \right.
\]

\[
\exp \left( \frac{\mu}{\epsilon} \sum_{\ell=0}^{k-2} L(x_{\ell}, u_{\ell}) \right) Z_{k-1} \bigg| Y_k \bigg] = \bar{E} \left[ \int_{\mathbb{R}^N} \eta (a(x_{k-1}, u_{k-1}) + w) \exp \left( \frac{\mu}{\epsilon} L(x_{k-1}, u_{k-1}) \right) \Psi(x_{k-1}, u_k) \right.
\]

8
\[
\exp \left( \mu \sum_{\ell=0}^{k-2} L(x_\ell, u_\ell) \right) Z_{k-1} \psi(w) dw \bigg| \gamma_k \right]
\]
\[
= \left\langle \sigma_{k-1}^{\rho}, \int_{\mathbb{R}^N} \eta(., u_{k-1}) + w \exp \left( \frac{\mu}{\epsilon} L(., u_{k-1}) \right) \Psi(., u_k) \psi(w) dw \right\rangle
\]
\[
= \langle \sigma_{k-1}^{\rho}, \Sigma(u_{k-1}, y_k) \eta \rangle
\]
\[
= \langle \Sigma^*(u_{k-1}, y_k) \sigma_{k-1}^{\rho}, \eta \rangle
\] (8)

This holds for all test functions \( \eta \in L^\infty(\mathbb{R}^N) \) hence the recursion result follows.

The fact that \( \Sigma^* \) maps \( L^1(\mathbb{R}^N) \) in \( L^1(\mathbb{R}^N) \) follows from (6) and the assumed properties of \( \psi, \Psi \) and \( L \). \( \square \)

The operator \( \Sigma \) actually maps \( C_b(\mathbb{R}^N) \) into \( C_b(\mathbb{R}^N) \) and hence we can define processes \( \nu_k(\tau) \in C_b(\mathbb{R}^N) \) for each \( \tau \in [0, T] \) by

\[
\begin{cases}
\nu_{k-1}(\tau) = \Sigma(u_{k-1}, y_k) \nu_k(\tau) & \text{for } 0 \leq k < \tau \\
\nu_{\tau}(\tau) = \exp \left( \frac{\mu}{\epsilon} \Phi \right) & \text{for } k = \tau.
\end{cases}
\] (9)

From the definition of the adjunct, (7) and (9) it is straightforward to establish the adjoint relationships
\[ \langle \sigma_k, \nu_k(\tau) \rangle = \langle \sigma_{k-1}, \nu_{k-1}(\tau) \rangle \] for \( 1 \leq k \leq \tau \).

2.4 Alternative Representation of the Cost

Let us introduce the following alternative representation of the cost associated with the new process \( \sigma_k \):

\[ K(u, \tau) = \bar{E} \left[ \langle \sigma_{\tau}, \exp \left( \frac{\mu}{\epsilon} \Phi \right) \right] \right]. \] (10)

**Theorem 3** We have that for all \( \tau \in [0, T] \) and \( u \in \mathcal{U}_{0, \tau-1} \)

\[ J(u, \tau) = K(u, \tau) \] (11)
**PROOF.** By (7) and (3) and conditional expectation properties [6, Page 331]. □

**Remark 4** The partially observed optimal stopping problem has now be rewritten as a fully observed (in state $\sigma_k$) but infinite dimensional optimal stopping control problem. The dynamics of the information state are governed by (7) and the cost given by (10).

**Remark 5** Note that the forward information state for this partially observed optimal stopping risk-sensitive control problem has the same structure as for the finite horizon version of this problem; however, an adjunct information state process is required for each of the possible stopping times.

### 3 Dynamic Programming for Risk-sensitive Optimal Stopping

The alternative partial observed optimal stopping control problem can now be solved using dynamic programming. Consider the state $\sigma_k$ on the interval $k, \ldots, \tau$ for $\tau \leq T$ with initial condition $\sigma_k = \sigma \in L^1(R^N)$:

\[
\begin{align*}
\sigma_{\ell} &= \Sigma^*(u_{\ell-1}, y_\ell)\sigma_{\ell-1} \quad \text{for } k + 1 \leq \ell \leq \tau \\
\sigma_k &= \sigma.
\end{align*}
\]

(12)

The corresponding value function for this problem is defined for $\sigma \in L^1(R^N)$ and $k \in [0, T]$ by

\[
S(\sigma, k) = \min_{\tau \in [k, T]} \inf_{u \in \mathcal{U}_{k, \tau-1}} \mathbb{E} \left[ \langle \sigma_k, \nu_k(\tau) \rangle | \sigma_k = \sigma \right].
\]

(13)
Theorem 6  The value function \( S(\sigma, k) \) satisfies the (backward) dynamic programming recursion

\[
\begin{aligned}
S(\sigma, k) &= \min \left( \langle \sigma, \exp \frac{\mu}{\epsilon} \Phi \rangle, \inf_{u_k \in U} \bar{E} \left[ S(\Sigma^*(u_k, y_{k+1})\sigma, k + 1) \right] \right) \\
S(\sigma, T) &= \langle \sigma, \exp \frac{\mu}{\epsilon} \Phi \rangle.
\end{aligned}
\]  

\( (14) \)

PROOF. Let

\[
\bar{S}(\sigma, k, \tau) = \inf_{u \in U_{k, \tau-1}} \bar{E} \left[ \langle \sigma_k, \nu_k(\tau) \rangle | \sigma_k = \sigma \right]
\]

\( (15) \)

defined for \( k \leq \tau \). We notice that \( \bar{S}(\sigma, k, \tau) \) is value function for the partially observed finite horizon risk-sensitive stochastic control problem with horizon \( \tau \).

Hence, it follows from the dynamic programming result established in [8] that \( \bar{S}(\sigma, k, \tau) \) is given by the following dynamic programming equation

\[
\bar{S}(\sigma, k, \tau) = \begin{cases} 
\langle \sigma, \exp \frac{\mu}{\epsilon} \Phi \rangle & \text{if } k = \tau \\
\inf_{u_k \in U} \bar{E} \left[ \bar{S}(\Sigma^*(u_k, y_{k+1})\sigma, k + 1, \tau) \right] & \text{if } k < \tau
\end{cases}
\]

Now noting that

\[
S(\sigma, k) = \min_{\tau \in [k, T]} \bar{S}(\sigma, k, \tau)
\]

\( (16) \)

and considering the \( \tau = k \) possibility directly we have

\[
S(\sigma, k) = \min \left( \langle \sigma, \exp \frac{\mu}{\epsilon} \Phi \rangle, \min_{\tau \in [k+1, T]} \inf_{u_k \in U} \bar{E} \left[ \bar{S}(\Sigma^*(u_k, y_{k+1})\sigma, k + 1, \tau) \right] \right).
\]

Then using (15) on the second term in the minimisation we obtain
\[
S(\sigma, k) = \min \left( \langle \sigma, \exp \frac{\mu}{\epsilon} \Phi \rangle, \min_{\tau \in [k+1, T]} \inf_{u_k \in U} \bar{E} \left[ \inf_{u_k+1, \tau-1} \bar{E} \left[ \langle \Sigma^*(u_k, y_{k+1}) \sigma, \nu_{k+1}(\tau) \rangle \right] \right] \right).
\]

Here the order of the infimum and minimum operations can be interchanged as can the order of the expectation and minimum operations (from the lattice property of \(\tau\) [5, Lemma 16.11]).

Hence, using (13) we have

\[
S(\sigma, k) = \min \left( \langle \sigma, \exp \frac{\mu}{\epsilon} \Phi \rangle, \inf_{u_k \in U} \bar{E} \left[ \min_{\tau \in [k+1, T]} \inf_{u_k+1, \tau-1} \bar{E} \left[ \langle \Sigma^*(u_k, y_{k+1}) \sigma, \nu_{k+1}(\tau) \rangle \right] \right] \right) \text{ as required.} \quad \Box
\]

3.1 Verification

**Theorem 7** Let \(\sigma'^*_k\) denote the information state sequence given by (7) initialised by \(\sigma'^*_0 = \rho\). Suppose that \(\tau^* \in [0, T]\) is the smallest time index \(k\) such that \(S(\sigma'^*_k, k) = \langle \sigma'^*_k, \exp \frac{\mu}{\epsilon} \Phi \rangle\) and further suppose that \(u^* \in U_0, \tau^*-1\) is a control policy so that, for each \(k = 0, \ldots, \tau^*-1\), \(u^*_k = \bar{u}_k(\sigma'^*_k)\) where \(\bar{u}_k(\sigma)\) achieves the minimum in (14). Then \(\tau^* \in [0, T]\) is the optimal stopping time and \(u^* \in U_0, \tau^*-1\) is the optimal control policy for the partially observed optimal stopping risk-sensitive stochastic control problem.

**Proof.** Parts of this proof follow the ideas used in the verification lemma presented in [8,7].
Define for \( \tau \in [0, T] \) and \( u \in U_{0, \tau - 1} \)

\[
\tilde{S}(\rho, k; u, \tau) = \begin{cases} 
\bar{E} [\langle \sigma_k^u, \nu_k(\tau) \rangle | \rho, u, \tau] & \text{for } k < \tau \\
\langle \sigma_k^u, \exp \frac{\epsilon}{\tau} \Phi \rangle & \text{for } k \geq \tau 
\end{cases}
\]

We will claim that

\[ S(\sigma_k^u, k) = \tilde{S}(\rho, k; u^*, \tau^*) \quad (17) \]

for each \( k = 0, 1, \ldots, T \).

For \( k = T \), (17) holds because \( S(\sigma_T^u, T) = \tilde{S}(\rho, T, u^*, \tau^*) = \langle \sigma_T^u, \exp \frac{\epsilon}{\tau} \Phi \rangle \). Now assume that (17) holds for \( k + 1, \ldots, T \). Consider now time index \( k \). From definition,

\[
\tilde{S}(\rho, k; u^*, \tau^*) = \begin{cases} 
\bar{E} [\langle \sigma_k^u, \nu_k(\tau^*) \rangle | \rho, u^*, \tau^*] & \text{for } k < \tau^* \\
\langle \sigma_k^u, \exp \frac{\epsilon}{\tau^*} \Phi \rangle & \text{for } k \geq \tau^* 
\end{cases}
\]

Using the same steps given in [8] for \( k < \tau^* \) we obtain

\[ \tilde{S}(\rho, k; u^*, \tau^*) = \begin{cases} 
\bar{E} [S(\Sigma(u_k, y_{k+1}) \sigma, k + 1)] = \bar{E}[S(\sigma_{k+1}^u, k + 1)] & \text{for } k < \tau^* \\
\langle \sigma_k^u, \exp \frac{\epsilon}{\tau^*} \Phi \rangle & \text{for } k \geq \tau^* 
\end{cases} \quad (18) \]

From the definition of \( \tau^* \) we have that \( \inf_{u_k \in U} \bar{E} [S(\Sigma(u_k, y_{k+1}) \sigma_k^u, k + 1)] < \langle \sigma_k^u, \exp \frac{\epsilon}{\tau^*} \Phi \rangle \) for all \( k < \tau^* \). Further, from assuming (17) holds for \( k + 1, \ldots, T \) we have that \( \bar{E}[S(\sigma_{k+n}^u, k+n)] \geq \langle \sigma_{k+n}^u, \exp \frac{\epsilon}{\tau^*} \Phi \rangle \) for \( n = \max(1, \tau^* - k), \ldots, T - k \).

Hence, (18) can be rewritten as follows:
\[
\bar{S}(\rho; k; u^*, \tau^*) = \min \left( \langle \sigma^g_k, \exp \frac{\mu}{\epsilon} \Phi \rangle, E \left[ S(\Sigma(u^*_k, y_{k+1})\sigma^g_k, k + 1) \right] \right)
\]

\[
= S(\sigma^g_k, k) \text{ for } k \in [0, T]
\]

from (14). By induction this proves (17).

Now from (17) and setting \( k = 0 \) we have

\[
\bar{S}(\rho, 0; u^*, \tau^*) = S(\rho, 0) \leq \bar{S}(\rho, 0, u, \tau)
\]

for any \( \tau \in [0, T] \) and \( u \in U_{0,\tau-1} \). Comparing (10) with definitions of \( \bar{S} \) and \( S \) we have that \( K(u^*, \tau^*) \leq K(u, \tau) \) for all pairs of \( \tau \in [0, T] \) and \( u \in U_{0,\tau-1} \) and this completes the proof. \( \square \)

Remark 8 The significance of Theorem 7 is that it establishes that the optimal policy for the partially observed optimal stopping risk-sensitive stochastic control problem is separated through the information state process \( \sigma_k \) (a sufficient statistic for the optimal stopping risk-sensitive problem).

Remark 9 However, it should be noted that the dynamic programming equations are recursions in the infinite dimensional \( \sigma_k \) and it is not clear how to best achieve a numeric approximation to this recursion.

One initial suggested path for numeric approximation (without proof), is to first approximate the forward infinite dimensional information state equation by a suitable finite dimensional approximation (perhaps an extended Kalman filter in nearly linear problems). Then, use these statistics as a new state in a new (modified) control problem. This modified problem is now a fully observed finite-dimensional optimal stopping control problem for which the Markov chain approximation technique (see [15]) can be used to achieve an approximation of the backward control dynamic programming equation.
It would be hoped that the validity of this type of outlined approach would be
determined by the fidelity of the forward information state equation approxima-
tion, and that as the approximation fidelity is improved, the error introduced
would decrease.

3.2 The Optimal Stopping Time

The optimal stopping rule for this optimal stopping time problem can be
expressed as

\[
\begin{align*}
&\text{if } S(\sigma, k) = \langle \sigma, \exp \frac{\epsilon}{\xi} \Phi \rangle \text{ stop} \\
&\text{if } S(\sigma, k) < \langle \sigma, \exp \frac{\epsilon}{\xi} \Phi \rangle \text{ continue}
\end{align*}
\]  

This can be interpreted to mean that it is optimal to stop when the future
cost of continuing from \( \sigma \) is equal to (or greater than) the cost of stopping at
\( \sigma \).

4 The Risk-Neutral Control Problem

In this section we consider the special case of the risk-sensitive control problem
that occurs as the risk-sensitive parameter \( \mu \) tends to zero.

Define the bounded linear operator \( \Sigma^0_{*} : L^1(\mathbb{R}^N) \rightarrow L^1(\mathbb{R}^N) \) by

\[
\Sigma^0_{*}(u, y)\sigma(z) = \int_{\mathbb{R}^N} \psi(z - a(\xi, u))\Psi(\xi, y)\sigma(\xi)d\xi
\]  

(20)
Theorem 10  We have

$$\lim_{\mu \to 0} \Sigma^\mu(u, y)\sigma = \Sigma^0(u, y)\sigma$$

uniformly on bounded sets of $U \times R \times L^1(R^N)$.

PROOF. Follows from (6) and (20). \qed

Next we define a process $\sigma^0_k \in L^1(R^N)$ by the recursion

$$\begin{cases}
\sigma^0_k = \Sigma^0(u_{k-1}, y_k)\sigma^0_{k-1} & \text{for } k \leq \tau \\
\sigma^0_0 = \rho.
\end{cases}$$

We note that $\langle \sigma^0_k, . \rangle$ is given by $\langle \sigma^0_k, \eta \rangle = \bar{E}[\eta(x_k)|\mathcal{Y}_k]$.

Remark 11  The process $\sigma^0_k$ is an unnormalised conditional density of $x_k$ given $\mathcal{Y}_k$ and (22) is known as the Duncan-Mortensen-Zakai equation.

4.1 A Risk-Neutral Control Problem

We again consider the discrete-time stochastic system (1) and formulate a partially observed optimal stopping risk-neutral stochastic control problem with cost

$$J^0(u, \tau) = E \left[ \sum_{\ell=0}^\tau L(x_\ell, u_\ell) + \Phi(x_\tau) \right]$$

defined for $\tau \in [0, T]$ and $u \in \mathcal{U}_{0,\tau-1}$. 

16
Let us introduce a value function defined for \( \sigma \in L^1(R^N) \) as

\[
W(\sigma^0, k) = \min_{\tau \in [k, T]} \inf_{u \in U_{k, \tau-1}} \mathbb{E} \left[ \sum_{\ell=k}^{\tau-1} \langle \sigma^0_\ell, L(., u_\ell) \rangle + \langle \sigma^0_\tau, \Phi \rangle \bigg| \sigma^0_k = \sigma \right] \tag{24}
\]

We now establish the following result.

**Theorem 12** The unnormalised conditional density \( \sigma^0_k \) is an information state for the partially observed optimal stopping risk-neutral problem and the value function, \( W(\sigma^0, k) \), satisfies the following dynamic programming equation

\[
\begin{align*}
W(\sigma^0, k) &= \min \left( \langle \sigma^0, \Phi \rangle, \inf_{u \in U} \mathbb{E} \left[ \sum_{\ell=k}^{\tau-1} \langle \sigma^0_\ell, L(., u_\ell) \rangle + \langle \sigma^0_\tau, \Phi \rangle \bigg| \sigma^0_k = \sigma \right] + W(\Sigma^0(u_k, y_{k+1})\sigma^0, k + 1) \right) \\
W(\sigma^0, T) &= \langle \sigma^0, \Phi \rangle.
\end{align*}
\tag{25}
\]

**PROOF.** Let \( W(\sigma^0, k, \tau) = \inf_{u \in U_{k, \tau-1}} \mathbb{E} \left[ \sum_{\ell=k}^{\tau-1} \langle \sigma^0_\ell, L(., u_\ell) \rangle + \langle \sigma^0_\tau, \Phi \rangle \bigg| \sigma^0_k = \sigma \right] \).

Then the proof follows the same steps as the proof of Theorem 6 and the result is established. \( \square \)

**Remark 13** Theorem 12 is the stopping time version of the risk-neutral finite-horizon dynamic programming equations given in [7,8].

5 Example: Guidance

In [16], it is shown that the guidance problem can be posed as a fully observed optimal stopping control problem. The optimal stopping and control framework is natural for guidance problems because in most real situations there is a natural freedom to choose the time-horizon for the trajectory that best suit the performance objectives (restricting the time-horizon artificially...
constrains the problem). Here, we extend the ideas of [16] by considering a partially observed risk-neutral version of the guidance problem.

Consider the following discrete-time approximation for the dynamics describing guidance to a stationary target (see [16] for more details):

\[
\begin{align*}
    r_{k+1} &= r_k - \cos(\theta_k) \Delta T \\
    \theta_{k+1} &= \theta_k + \left[ u_k + \frac{1}{r_k} \sin(\theta_k) \right] \Delta T
\end{align*}
\]

where \( r_k \) and \( \theta_k \) are the range and the line-of-sight angle to the target at time instant \( k \), \( u_k \in [-U_{\text{max}}, U_{\text{max}}] \) is a control action, and \( \Delta T \) is some chosen sample period.

We assume noisy measurements of the range and bearing are available as follows:

\[
y_k = [r_k, \theta_k]' + w_k
\]

where \( w_k \in \mathbb{R}^2 \) is a vector of two mutually independent zero mean i.i.d. gaussian noises. Alternatively, the partially observed problem with only bearing angle information is another important problem.

Assuming a finite initial condition, and because the dynamics involve finite changes at each time step, \( r_k \) is bounded in the sense that \( |r_k| < R_{\text{max}} \) for \( k = 0, 1, \ldots, T \). Hence we can consider the non-negative bounded and continuous stopping and running costs

\[
\Psi = \begin{cases}
    |r| & \text{for } |r| < R_{\text{max}} \\
    R_{\text{max}} & \text{otherwise}
\end{cases} \quad \text{and} \quad L = \begin{cases}
    u^2 & \text{for } |u| < U_{\text{max}} \\
    U_{\text{max}}^2 & \text{otherwise}
\end{cases}.
\]
According to Theorem 12, the partially observed optimal stopping solution is described by the dynamic programming equation:

\[
\begin{align*}
W(\sigma^0, k) &= \min \left( \langle \sigma^0, \Phi \rangle, \inf_{u_k \in U} \left\{ u_k^2 + \bar{E} \left[ W(\Sigma^{*0}(u_k, y_{k+1})\sigma^0, k + 1) \right] \right\} \right) \\
W(\sigma^0, T) &= \langle \sigma^0, \Phi \rangle.
\end{align*}
\]

(26)

where we have used that \( \bar{E} [\langle \sigma^0, L(., u_k) \rangle] = u_k^2 \). Further, when coupled with the information state, we will have \( \langle \sigma_k^0, \Phi \rangle = \bar{E} \left[ |r_k| |\mathcal{Y}_k| \right] \) for all \( k = 0, 1, \ldots, T \).

From [16] we know that in the fully observed case, the optimal control and stopping rule depend on both range and bearing information; and (26) demonstrates that in the same occurs here (although through a more complicated dependence). For example, the nature of the infimum term in (26) highlights that the optimal control will depend on the whole information state, not just the conditional mean estimates of \( r \) and \( \theta \).

In general terms, (26) demonstrates that guidance based solely on measurement based state estimates is not optimal. However, in low noise situations, where the support of the information density will be small, the resulting performance the loss will be minimal (and this is consistent with the general practice of using only state estimates in guidance if good measurements are available).

If only noisy measurements are available then further development of techniques for solving this problem are required because the information state recursion (22) is infinite dimensional.

One possible method for developing approximation solutions for this problem is to approximate the information state recursion using an extended Kalman
filter recursion. Then the dynamic programing equation (26) could be rewritten in terms of the finite dimensional mean and variance information. This new problem is fully observable and finite dimensional and approximation solutions can be developed using the Markov chain approach given in [15]. Admittedly, there are currently no convergence results for this sort of approximation approach to the infinite dimensional partially observed problem.

6 Conclusions

In this paper we introduce the partially observed non-linear risk-sensitive optimal stopping control problem for non-linear discrete-time systems. We established a separation principle and dynamic programming equations through an infinite dimensional information state approach. Dynamic programming equations for the risk-neutral variant of the problem are also given. A simple example is used to illustrate some features of the problem.

References


